

Grado en Biotecnología  
Trabajo de Fin de Grado



# Estudio computacional de interacciones ligando-proteína en glicosiltransferasas mediante dinámica molecular

---

*(Computational study of ligand-protein interactions  
in glycosyltransferases through molecular dynamics)*

**Director: Dr. Pedro Merino Filella**  
Departamento de Química Orgánica, Facultad de Ciencias

**David Sánchez Navarro**

Curso 2020/2021

# Índice

<b>Resumen .....</b>	<b>2</b>
<b>Abstract.....</b>	<b>3</b>
<b>1. Introducción .....</b>	<b>4</b>
<b>2. Antecedentes .....</b>	<b>6</b>
<b>3. Objetivos .....</b>	<b>7</b>
<b>4. Métodos computacionales .....</b>	<b>8</b>
<b>4.1. Preparación del archivo PDB .....</b>	<b>8</b>
4.1.1. Modelado de residuos ausentes .....	8
4.1.2. Caracterización de histidinas y cisteínas .....	9
4.1.3. Caracterización de ligandos .....	9
<b>4.2. Dinámica molecular .....</b>	<b>10</b>
4.2.1. Obtención de parámetros .....	10
4.2.2. Lanzamiento de la simulación.....	10
<b>4.3. Visualización de resultados.....</b>	<b>11</b>
<b>5. Resultados y discusión.....</b>	<b>12</b>
<b>5.1. Cálculos de dinámica molecular .....</b>	<b>12</b>
<b>5.2. Comparación de POFUT1 humana y de <i>C. elegans</i> .....</b>	<b>13</b>
<b>5.3. Comparación de <i>Hs</i>POFUT1 en forma apo y con GDP-fucosa .....</b>	<b>15</b>
<b>5.4. Estudio de interacciones con el ligando .....</b>	<b>17</b>
<b>6. Conclusión .....</b>	<b>19</b>
<b>Referencias .....</b>	<b>20</b>

## Resumen.

### *“Estudio computacional de interacciones ligando-proteína en glicosiltransferasas mediante dinámica molecular”*

Las glicosiltransferasas son enzimas responsables de la biosíntesis y el procesamiento de glicanos en macromoléculas. Hay una gran cantidad y variedad de enzimas que se incluyen en esta clasificación, entre ellas la POFUT1, que cataliza la transferencia de un residuo de fucosa desde el dinucleótido GDP-fucosa a proteínas implicadas en la vía de señalización Notch, un mecanismo de comunicación celular cuya desregulación se relaciona con diferentes patologías como la leucemia o el cáncer.

En este trabajo se procede a un análisis de la proteína POFUT1 mediante simulaciones de dinámica molecular para hacer cálculos de estabilidad, estudios comparativos de la proteína en diferentes organismos y distintas conformaciones, y finalmente un estudio de las interacciones que establece el centro activo con el ligando GDP-fucosa, con el objetivo de establecer predicciones que puedan servir de base para futuros experimentos.

## **Abstract.**

### *“Computational study of ligand-protein interactions in glycosyltransferases through molecular dynamics”*

Glycosyltransferases are enzymes responsible for the biosynthesis and processing of glycans in macromolecules. There's a significant amount and variety of these enzymes, among which is POFUT1, responsible of transferring a fucose unit from the dinucleotide GDP-fucose to proteins involved in the Notch signaling pathway. A lack of regulation in this cell communication system is the cause of many health problems such as leukemia or certain types of cancer.

The purpose of this work is to perform an analysis of protein POFUT1 using molecular dynamics simulations, in order to carry out stability calculations, comparative studies in different organisms and conformational states, and finally a study of the active site interactions with its ligand, GDP-fucose, completing a report that can be used as a basis for future experiments.

## 1. Introducción.

Las glicosiltransferasas son enzimas responsables de la biosíntesis y el procesamiento de glicanos en macromoléculas. Hay una gran cantidad y variedad de enzimas que se incluyen en esta clasificación, que difieren en el monosacárido que es transferido, generalmente a partir de un azúcar unido a un nucleótido (GDP-fucosa, UDP-galactosa); y el sustrato sobre el que actúan, que no solo incluye proteínas, sino también DNA o incluso lípidos.<sup>1</sup>

La mayoría de estas glicosiltransferasas intervienen en la glicosilación de péptidos y proteínas y en la elongación de los glicanos, en este caso añadiendo azúcares sobre unidades ya existentes, mientras que el resto se encargan de la transferencia del primer azúcar, como la enzima POFUT1, cuyo estudio computacional ha sido el objeto de estudio de este trabajo.<sup>2</sup>

La POFUT1 (*protein O-fucosyltransferase 1*) es una enzima perteneciente a la familia de las O-fucosiltransferasas. Cataliza la adición de una fucosa, a partir de una molécula de GDP-fucosa, mediante un enlace O-glicosídico a un residuo conservado de serina o treonina en unas secuencias ricas en cisteína conocidas como repeticiones EGF-like (*epidermal growth factor-like*), presentes en determinadas proteínas.<sup>3</sup>

Entre los sustratos que son susceptibles a la fucosilación por POFUT1 cobran gran relevancia las proteínas de la vía de señalización Notch, un mecanismo de comunicación celular que determina el destino y diferenciación celular. Este sistema de transducción de señales comprende 4 proteínas Notch a modo de receptores (Notch-1 a 4, diferentes en el número de repeticiones EGF-like) y 5 ligandos Notch (Jagged-1 y 2, DLL-1, 3 y 4).<sup>4</sup> La glicosilación de estas proteínas es esencial para su correcto funcionamiento.

La fucosilación iniciada por POFUT1 continúa con la acción de una glicosiltransferasa denominada Fringe, que añade una N-acetilglucosamina sobre la fucosa, y posteriormente una galactosiltransferasa y una sialiltransferasa que añaden una galactosa y un ácido siálico, respectivamente, para completar un tetrasacárido.<sup>5</sup> Otras glicosiltransferasas que actúan sobre las proteínas Notch son POGLUT1 (*protein O-glucosyltransferase 1*) o EOGT (*EGF domain specific O-linked N-acetylglucosamine transferase*), como queda representado en la FIGURA 1.<sup>6</sup>

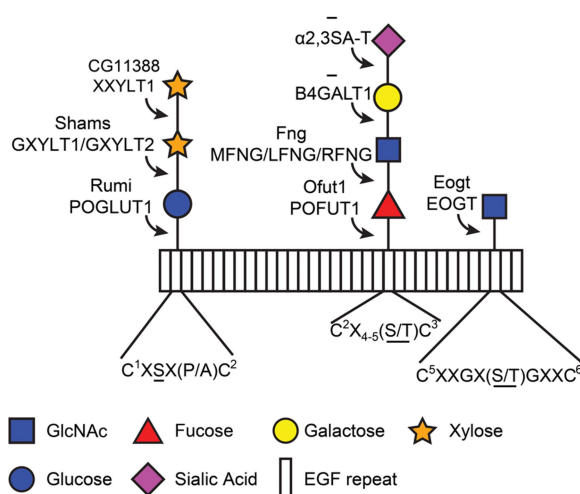


FIGURA 1. Representación esquemática de los glicanos en las repeticiones EGF-like de las proteínas Notch, indicando los azúcares que los forman y las enzimas que catalizan su adición.<sup>6</sup>

Tanto los receptores como los ligandos Notch son proteínas transmembrana expuestos en la superficie de las células. La unión del ligando al receptor Notch induce el corte y la liberación del dominio intracelular de Notch (NICD) por una serie de proteasas, de las cuales es relevante la  $\gamma$ -secretasa en la clínica (se está estudiando el desarrollo de inhibidores químicos en el tratamiento del cáncer<sup>7</sup>). El NICD se transloca al núcleo donde activa factores de transcripción y activa la expresión de determinados genes (FIGURA 2).<sup>4</sup>

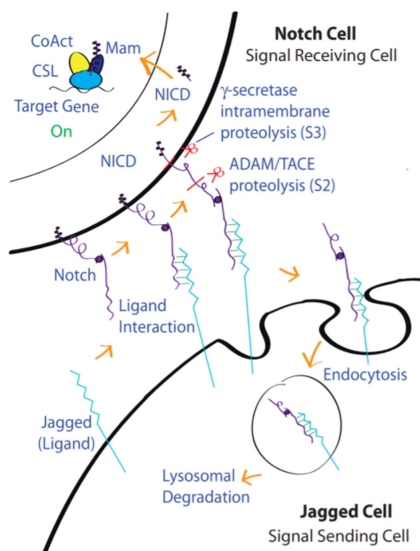


FIGURA 2. Proceso de transducción de señales en la vía de señalización Notch.<sup>4</sup>

Cabe destacar la importancia de la vía Notch por ejemplo en linfocitos T, donde regula la diferenciación  $\alpha\beta/\gamma\delta$  o CD4/CD8 y la proliferación y supervivencia de estas células, y fallos en este proceso pueden resultar en leucemia.<sup>8</sup>

La vía de señalización está también implicada en muchos tipos de cáncer porque, al estar encargada del mantenimiento y desarrollo de los tejidos, una falta de regulación que provoca un exceso de señal contribuye al crecimiento de tumores<sup>9</sup>. Otras complicaciones derivadas de fallos en la señalización Notch son esclerosis múltiple o tetralogía de Fallot (defecto cardíaco por anomalías en la homeostasis de las válvulas).<sup>10</sup>

Se está estudiando la manera de bloquear o atenuar la vía Notch para el tratamiento de estas patologías, y una de las estrategias que se plantea es el desarrollo de inhibidores contra POFUT1. Por eso cobra gran relevancia el estudio computacional de esta proteína, prestando especial atención a su ligando GDP-fucosa, ya que análogos estructurales con sustituyentes de diferente naturaleza en lugar de la fucosa ofrecen resultados prometedores como inhibidores.<sup>11</sup> Algunos de los inhibidores de glicosiltransferasas (no para POFUT1) descritos hasta la fecha se muestran en la FIGURA 3:

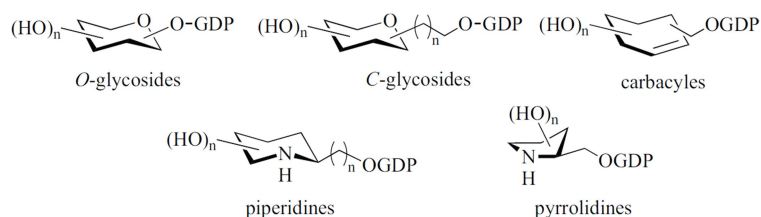


FIGURA 3. Inhibidores de glicosiltransferasas.<sup>11</sup>

## 2. Antecedentes.

Este Trabajo de Fin de Grado ha sido realizado bajo la supervisión del Prof. Pedro Merino, catedrático del departamento de Química Orgánica, en la Facultad de Ciencias de la Universidad de Zaragoza y miembro del Instituto de Biocomputación y Física de Sistemas Complejos (BIFI). Entre las líneas de investigación del instituto se encuentra el campo de la glicobiología, profundizando en el estudio de glicosiltransferasas y transglicosilasas.<sup>12</sup>

De las glicosiltransferasas se estudian concretamente aquellas que utilizan dadores de Leloir (mono o difosfonucleótidos con un azúcar) derivados de uridina (por ejemplo, OGT emplea UDP-GlcNAc y GalNAc-T2 usa UDP-GalNAc) o guanosina (tanto POFUT1 como POFUT2 utilizan GDP-fucosa). Entre ellas destacan las representadas en la FIGURA 4:

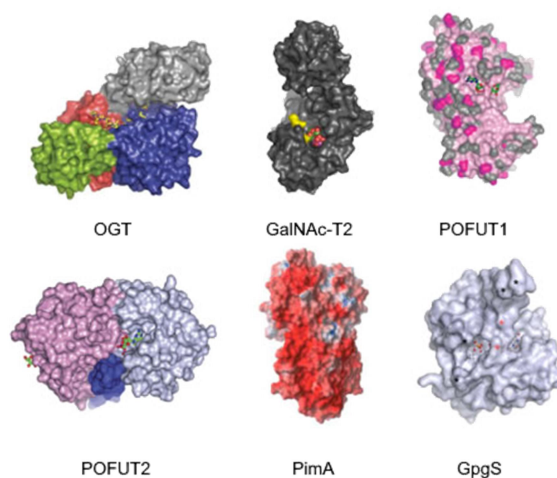


FIGURA 4. Ejemplos de enzimas estudiadas en el grupo de investigación del departamento.<sup>12</sup>

Respecto a POFUT1, la primera estructura cristalizada resuelta de esta proteína se atribuye a este grupo de investigación<sup>13</sup>, y dos de las estructuras descritas se muestran en la FIGURA 5. Corresponde al organismo *Caenorhabditis elegans* ("CePOFUT1"), y presenta una similitud del 41% con sus homólogos en otras especies eucariotas, aunque no tiene sitios de N-glicosilación.

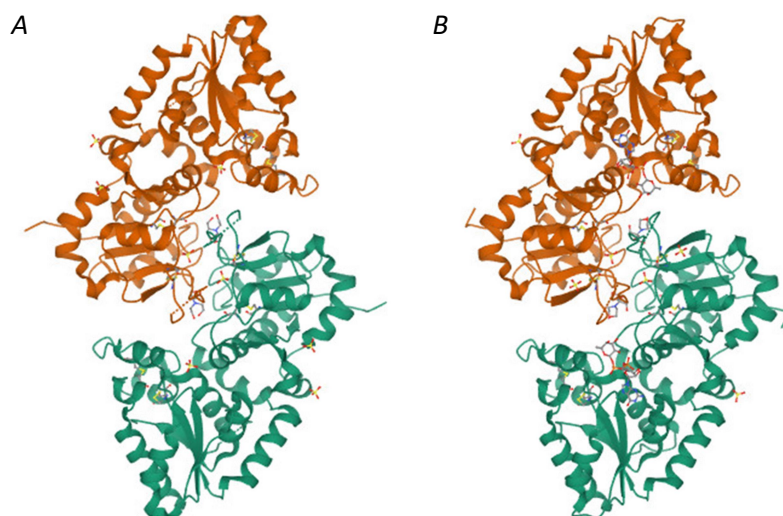


FIGURA 5. A: Estructura cristalizada de CePOFUT1 en su forma apo (ID PDB: 3ZY4).<sup>14</sup>  
B: Estructura cristalizada de CePOFUT1 en complejo con GDP-fucosa (ID PDB: 3ZY5).<sup>15</sup>

Este TFG también se nutre de los estudios computacionales del instituto, aplicando el enfoque de las simulaciones de dinámica molecular al estudio de las estructuras disponibles experimentalmente.

En la química computacional se distingue entre los cálculos de MD (dinámica molecular utilizando parámetros de mecánica molecular, MM), para describir mecanismos biológicos como la minimización de estructuras de proteínas o la interacción con ligandos para determinar sitios de unión; y QM (mecánica cuántica), que requieren mayor poder de computación y son ideales para simular mecanismos de reacciones químicas.

### 3. Objetivos.

El objetivo global de este trabajo es el empleo de simulaciones de dinámica molecular para evaluar la estabilidad en la estructura de POFUT1 y el posterior uso de herramientas bioinformáticas para hacer un análisis comparativo y estructural de la proteína de diversas procedencias.

Para la realización de este objetivo se ha partido de tres estructuras cristalizadas ya resueltas de la proteína.

La primera de ellas es la POFUT1 del organismo *C. elegans* “CePOFUT1” (ID PDB: 3ZY4<sup>14</sup>), en este caso la apoproteína (no tiene ligando). Las otras dos corresponden a la proteína en *Homo sapiens* “HsPOFUT1”, tanto la estructura en forma apo (ID PDB: 5UX6<sup>16</sup>) como unida al ligando GDP-fucosa (ID PDB: 5UXH<sup>17</sup>).

En base a estas estructuras y simulaciones se definen una serie de objetivos secundarios.

- Cálculos de dinámica molecular de las proteínas estudiadas.
- Estudio comparativo de la proteína CePOFUT1 con HsPOFUT1.
- Comparación de las estructuras de las proteínas humanas en forma apo y con ligando.
- Evaluación de interacciones con el ligando y comparación con valores experimentales.



## 4. Métodos computacionales.

Todas las simulaciones de dinámica molecular se llevaron a cabo en equipos propios del grupo de investigación equipados con tarjetas GPU sobre las que se ejecutó el software AMBER<sup>18</sup> en un entorno Linux. El análisis de los resultados se realizó con los programas integrados en el paquete AmberTools. Para la preparación se utilizó el programa Chimera<sup>19</sup> y para la visualización el programa PyMol<sup>20</sup>.

### 4.1. Preparación del archivo PDB.

Para poder lanzar la simulación hay que disponer del archivo PDB que contiene la estructura tridimensional de la proteína, el cual hay que corregir eliminando las moléculas de agua y otras no relevantes que aparecen por ser residuos de la cristalización (SO<sub>4</sub>, BTB). También será necesario hacer otras modificaciones en el PDB que se detallan a continuación.

#### 4.1.1. Modelado de residuos ausentes.

Comparando la estructura disponible con la información proporcionada en la literatura, se observa que faltan varios residuos que estarían formando bucles, y que aparecen como líneas discontinuas al visualizar el PDB. Para modelar estos residuos ausentes se ha utilizado el software Chimera, que tiene una opción para modelar los bucles. El resultado final se muestra en la FIGURA 6:

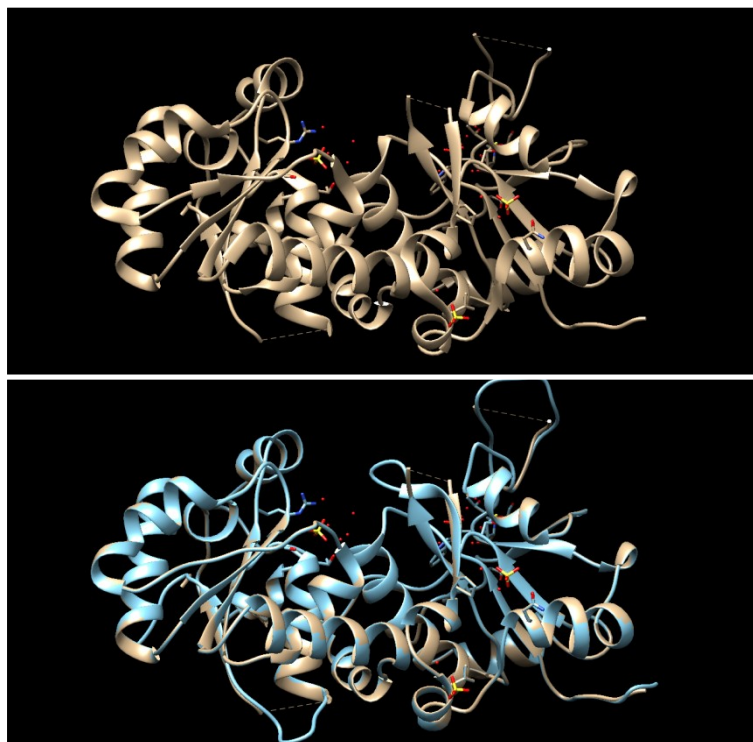


FIGURA 6. Estructura de CePOFUT1 antes (arriba) y después (abajo) de modelar los bucles. En color beige se muestra la estructura original y en azul la modelada por Chimera.

Hay que tener en cuenta que el archivo PDB que proporciona Chimera altera la numeración original asignando como residuo 1 al primer residuo de la estructura, que en el ejemplo de CePOFUT1 corresponde en realidad con el residuo número 22 en el PDB de partida.

#### 4.1.2. Caracterización de histidinas y cisteínas.

Es necesario también caracterizar las histidinas según el estado de protonación, e identificar las cisteínas que establecen puentes disulfuro. Las histidinas pueden estar en tres formas según su estado de protonación. Según se encuentren en la estructura original será necesario identificar el correspondiente tautómero en el fichero PDB de acuerdo con la nomenclatura indicada en la FIGURA 7.

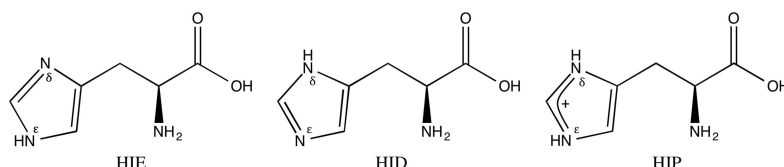


FIGURA 7. Diferentes estados de protonación del anillo de imidazol en la histidina.<sup>21</sup>

Las cisteínas que forman puentes disulfuro tendrán que ser identificadas en el PDB como CYX, el resto como CYS. La identificación de los aminoácidos mencionados puede hacerse utilizando el programa PyMol como se muestra en la FIGURA 8.

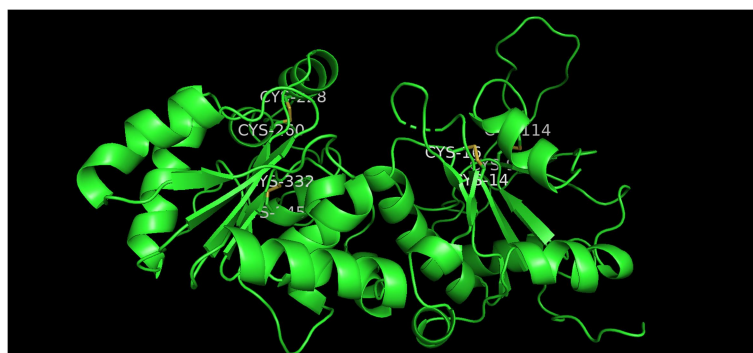


FIGURA 8. Estructura de CePOFUT1 vista en PyMol donde aparecen resaltados en color amarillo los puentes disulfuro y los átomos que los forman etiquetados mostrando su numeración.

#### 4.1.3. Caracterización de ligandos.

En el caso de la CePOFUT1 o la apo-HsPOFUT1 no es necesario hacer más cambios, sin embargo la HsPOFUT1 con GDP-fucosa presenta un ligando que hay que tener en cuenta a la hora de preparar la simulación. Hay que disponer del archivo PDB del ligando, que se puede extraer del PDB de la proteína cristalizada o a partir del código SMILES desde cualquier software de dibujo de estructuras químicas (ChemDraw, BioDraw, etc.) (FIGURA 9).

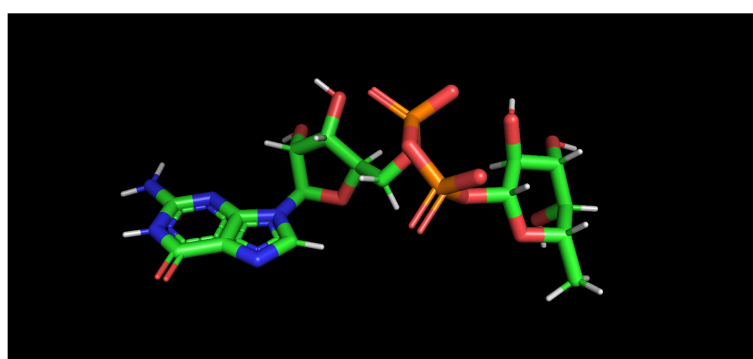


FIGURA 9. Representación 3D de la GDP-fucosa (GFB) visualizada en PyMol.

Mediante el programa Antechamber se generan los archivos .prep, y .frcmod que contienen los parámetros del ligando a estudiar y que se incorporarán al campo de fuerzas empleado en la preparación de la proteína mediante el programa Xleap, como se indica a continuación.

## 4.2. Dinámica molecular.

### 4.2.1. Obtención de parámetros.

El primer paso para poder lanzar la simulación es generar los archivos de topología (.prmtop) y de coordenadas (.inpcrd) con el programa Xleap del paquete AMBER. Para la generación de dichos archivos se utilizan los campos de fuerzas correspondientes al agua y solventes, carbohidratos y proteínas.

Tras la incorporación de los campos de fuerzas se indican los puentes disulfuro, se neutraliza la proteína y se añade una caja de aguas de 12 Å.

En el caso de que la proteína tenga algún ligando que se haya tenido que reconocer por Antechamber, como el caso de la HsPOFUT1 con GDP-fucosa, hay que introducir también los correspondientes archivos de parámetros.

De esta forma se obtiene un archivo *cepofut1\_MD.pdb*, que es la proteína preparada para la dinámica como se observa en la FIGURA 10 a continuación, y los archivos *cepofut1.prmtop* de parámetros de topología y *cepofut1.inpcrd* de coordenadas.

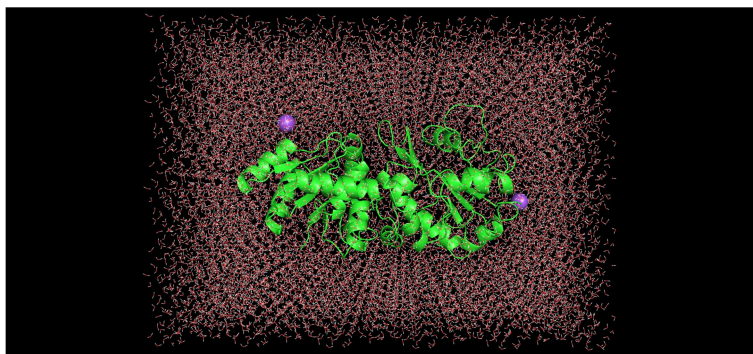


FIGURA 10. Archivo *cepofut1\_MD.pdb* visualizado con PyMol donde se aprecia la introducción de la caja de aguas y los iones  $\text{Na}^+$  añadidos por xleap.

### 4.2.2. Lanzamiento de la simulación.

Antes de lanzar la simulación son necesarios una serie de pasos previos:

- 1. Minimización del sistema manteniendo la proteína y el ligando fijos.
- 2. Minimización de todo el sistema.
- 3. Calentamiento del sistema hasta la temperatura deseada (300 K).
- 4. Equilibración del sistema a la temperatura alcanzada (300 K).

Una vez llevados a cabo estos cuatro pasos se lleva a cabo el cálculo definitivo de dinámica molecular de todo el sistema que proporcionará la evolución del sistema a lo largo del tiempo.

La dinámica molecular se calcula cada 2 femtosegundos ( $10^{-15}$  s) y los resultados se almacenan cada picosegundo ( $10^{-12}$  s). Posteriormente se visualizarán frames correspondientes a 1 nanosegundo ( $10^{-9}$  s) ya que en escalas menores de tiempo los cambios son imperceptibles.

### **4.3. Visualización de resultados.**

La visualización de la dinámica se realiza con el software PyMol extrayendo vistas correspondientes a cada nanosegundo de cálculo.

## 5. Resultados y discusión.

### 5.1. Cálculos de dinámica molecular.

Como resultado de los cálculos de dinámica molecular se han obtenido tres conjuntos de archivos PDB que constituyen los frames de la simulación para cada proteína analizada. La simulación de POFUT1 de *C. elegans* consta de 713 frames, y las de POFUT1 humana constan de 1000 frames (apo) y 136 frames (con ligando GDP-fucosa).

Cada frame se corresponde con un nanosegundo del proceso de dinámica molecular. Generalmente, las simulaciones se consideran exitosas al alcanzar los 500 nanosegundos, que es tiempo suficiente para los procesos a escala molecular, en este caso la estabilidad de las estructuras de proteínas.

Durante los cálculos para la HsPOFUT1 con GDP-fucosa ha habido ocasiones en las que había problemas con la dinámica, dejando unos valores de “NaN” en las coordenadas de los átomos de la proteína que hacen que se corte la simulación (en el primer intento se han podido rescatar 70 frames). Tras reequilibrar el sistema durante más tiempo y volver a lanzar la simulación, se han conseguido 136 frames antes de volver a desestabilizarse. En estos casos habría que empezar de cero y jugar con la estructura de la proteína, utilizando otra cadena o modificando los plegamientos ligeramente.

La herramienta principal que se utiliza para determinar la estabilidad de las estructuras es el cálculo de RMSD (*root-mean-square deviation*) a lo largo del proceso de dinámica, que mide la desviación respecto del archivo usado como referencia, en este caso el que se ha utilizado para lanzar la MD. Los valores de RMSD a lo largo del tiempo de la simulación de cada proteína se representan gráficamente en Excel, como se muestra en la FIGURA 11.

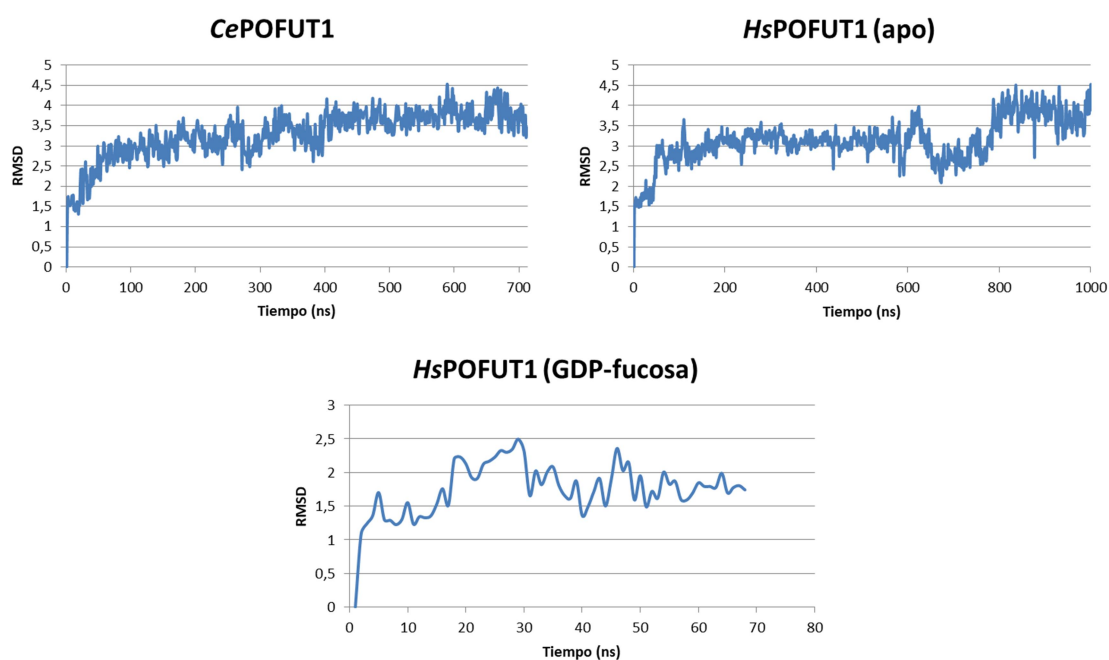
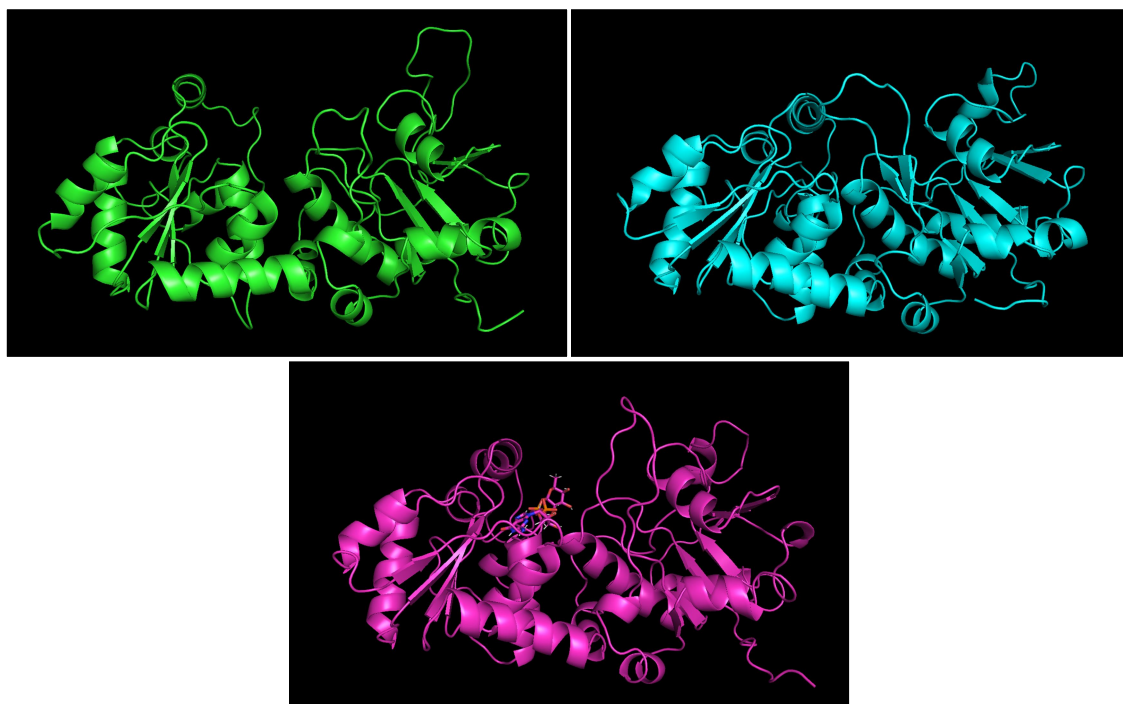


Figura 11. Representación gráfica de RMSD respecto al tiempo en las tres proteínas analizadas.

De estas gráficas se deduce la existencia de una estructura estable en los tres casos, que se muestra oscilando en un rango concreto de valores de RMSD que se mantiene relativamente constante en el tiempo. *CePOFUT1* se estabiliza a partir de los 400 ns, *apo-HsPOFUT1* en torno a 250 ns (hacia los 600 ns se desestabiliza otra vez), y *HsPOFUT1* con GDP-fucosa a los 30 ns, aunque no ha sido posible calcular más frames para asegurarlo.

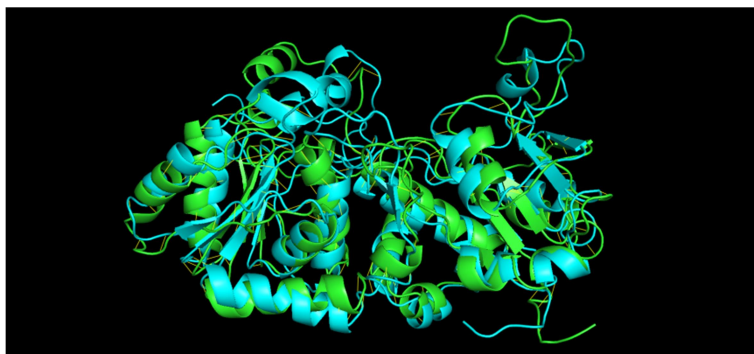
Los frames que he tomado en cada proteína para utilizar como estructuras estables en la resolución de los próximos objetivos son el 500 para *CePOFUT1*, el 500 para *HsPOFUT1* (apo) y el 50 para *HsPOFUT1* (con GDP-Fucosa), y se observan en la *FIGURA 12* a continuación.



*FIGURA 12. Estructuras estables de CePOFUT1 (verde), HsPOFUT1 en forma apo (azul) y HsPOFUT1 con GDP-fucosa (rosa) visualizadas en PyMol.*

## 5.2. Comparación de POFUT1 humana y de *C. elegans*.

Para poder comparar las estructuras de *CePOFUT1* y *HsPOFUT1* (apo), se han tomado los frames correspondientes a las estructuras estables, y en PyMol se ha hecho una superposición de las estructuras mediante un alineamiento. El resultado se muestra en la *FIGURA 13*.



*FIGURA 13. Superposición de las proteínas CePOFUT1 (verde) y apo-HsPOFUT1 (azul) en PyMol.*



El alineamiento presenta un RMSD de 2.460, y algunas diferencias estructurales notables que se observan son una disposición diferente en una de las  $\alpha$ -hélices (FIGURA 14A), o la conformación de segmentos con mucha libertad de movimiento como son los extremos N-terminales (FIGURA 14B) o la estructura de bucles (ejemplos en FIGURAS 14C y 14D).

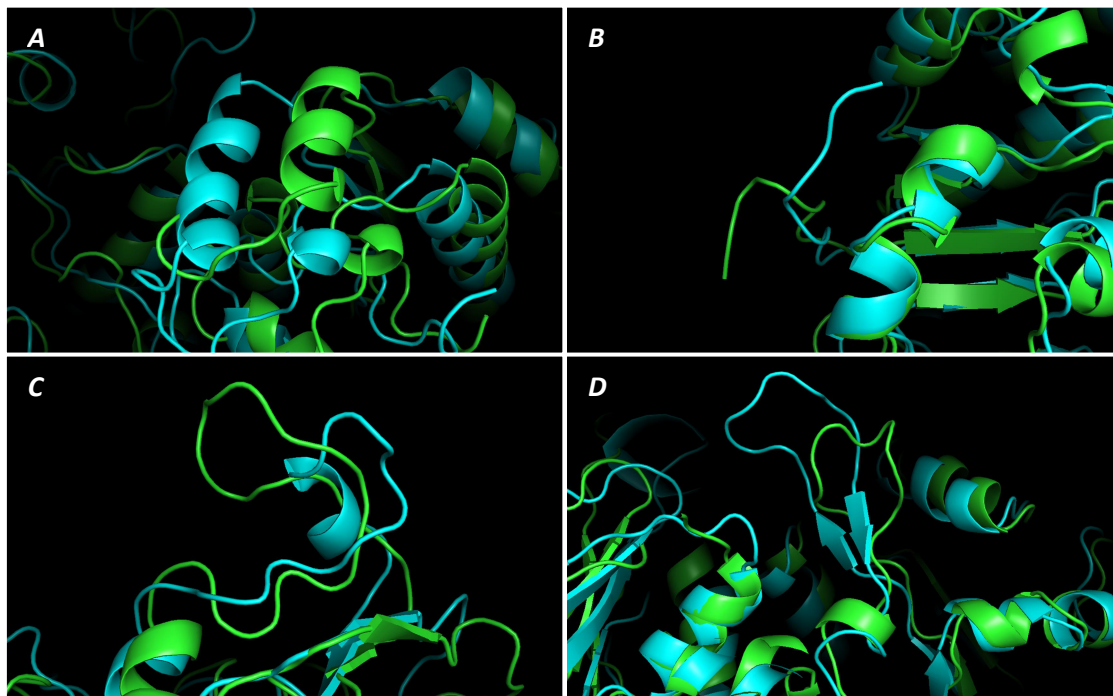


FIGURA 14. Diferencias en la  $\alpha$ -hélice (A. secuencia Ce: L218-A142, Hs: L216-A240); extremos N-terminales (B) y bucles (C. Ce: C98-F122, Hs: C103-F125; D. Ce: P44-F63, Hs: P45-F68).

Como las proteínas de ambas especies son diferentes en cuanto a su secuencia, mediante un alineamiento de las secuencias se podrán encontrar diferencias en los residuos que conforman el centro activo que podrían repercutir en la unión al ligando. Se ha utilizado la herramienta SIM (Expasy)<sup>22</sup> para hacer el alineamiento de secuencia, que se muestra en la FIGURA 15.

3ZY4(ce)	2	AEATDPNGYIVFCPCMG	RFGN	QVDQFLGVLAFAKALDRTLVLNPFIEFKH--PETKM-I
SUX6(hs)	3	AGSWDPAGYLLYCPCMG	RFGN	QADHFLGSLAFAKLLNRTLAVPPWIEYQHKKPPTNLHV
		* * * *		* * * * *
3ZY4(ce)	58	PFEFLFQVGTVAKYTRVVTMQEFTKKIMPTVWPPEKRKAFCHWTPRQAIYDKSAEPGCHSK		
SUX6(hs)	63	SYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPKRVAYCF--EVAAQRSPDKKTCRPMK		
		* * *		* * * * *
3ZY4(ce)	118	EGNPFPGPYWDQIDVSFVGDEYFGDIPGGFDLNQMSRKKWLEKFPSEEPVLAFFSSAPAP		
SUX6(hs)	121	EGNPFPGPYWDQFHVSAKSELFTGI--SFSASY--REQWSQRFSPKEHPVLALPGAPAQ		
		* * * *		* * * * *
3ZY4(ce)	178	FPSKGKVSIIQYLRWSSRITEQAKKFISANLAKPFVAV	HL	RNDADWVRVCEHI-DTTTN
SUX6(hs)	176	FPVLEEHRPLQKYMVWSDVMVKTGEAQIHAHLVRPYVGI	HL	IGSDWKNACAMLDGTAG
		* * *		* * * * *
3ZY4(ce)	237	RPL	ASEQCLGEGHHLGT-LTKEICSPSKQIQILEQIVEKVGSIKAKSVFVAS	SKDHMIDE
SUX6(hs)	236	SHF	ASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVAT	DSSEYVPE
		* * *		* * * * *
3ZY4(ce)	296	INEALKPYEIEAHRQEP	DDMYTSLAIMGRADLFVGNVST	FSHIVKRERDHAGQSPRPSA
SUX6(hs)	296	LQQLFKG-KVKVWSLKP	EVAQVDLYILGQADHFIGNCVS	STAFVKRERDLQG--RPSS
		* * *		* * * * *
3ZY4(ce)	356	FFGI		
SUX6(hs)	352	FFGM		
		***		

FIGURA 15. Alineamiento de secuencias donde se han coloreado los aminoácidos del centro activo: en verde los coincidentes y en rojo los discordantes.<sup>22</sup>

Atendiendo a los residuos de aminoácidos que conforman el centro activo de la proteína CePOFUT1 como se ha comprobado experimentalmente<sup>13</sup>, marcados en la FIGURA 13 con el correspondiente código de colores, hay cuatro de ellos que son distintos al observar la secuencia de la HsPOFUT1, y estos cambios podrían repercutir en las interacciones que establece con el ligando GDP-fucosa:

- La **fenilalanina-261** establece una interacción hidrofóbica de tipo stacking con la fucosa. Según el alineamiento cambia por metionina, que no tiene las mismas propiedades, pero el residuo anterior a esta es una **fenilalanina** por lo que este contacto podría estar conservado en una disposición conformacional diferente.
- El esqueleto de la **serina-308** establece un puente de hidrógeno con la guanina. Cambia por **treonina**, que añade un grupo metilo pero conserva el grupo hidroxilo que establece el puente, por lo que la interacción se mantendría.
- La cadena lateral del **ácido aspártico-334** establece un puente de hidrógeno con guanina. Cambia por un **ácido glutámico**, que mantiene las propiedades químicas pero cuenta con un grupo metileno adicional que hace que esta cadena sea más larga.
- La **treonina-356** establece un puente de hidrógeno con el fosfato- $\beta$ . Cambia por **serina**, que como se ha comentado antes se diferencia en un grupo metilo pero el grupo hidroxilo no cambia, por lo que el puente de hidrógeno se conservaría.

Para comprobar la conservación de estas interacciones se han visualizado, en las estructuras originales de CePOFUT1 y HsPOFUT1, los aminoácidos del sitio activo de ambas proteínas utilizando PyMol, como se representa en la FIGURA 15.

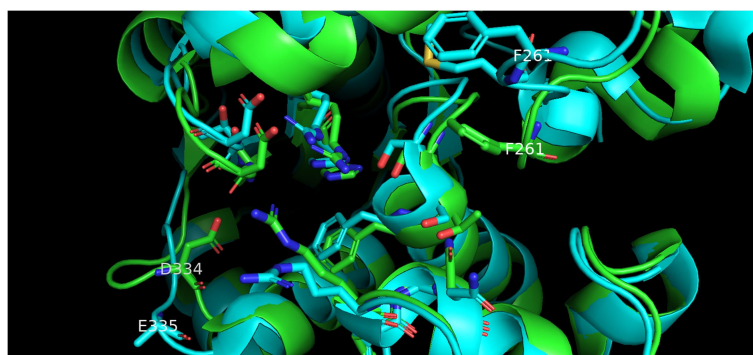


FIGURA 15. Visualización de los aminoácidos, con su cadena lateral, que comprenden el sitio activo de las proteínas CePOFUT1 (verde) y HsPOFUT1 (azul). Aparecen etiquetados los residuos más divergentes entre las dos estructuras.

Como cabe esperar, gran parte de estos residuos de aminoácidos presentan una disposición de su cadena lateral similar en ambas proteínas, demostrando que sus interacciones con el ligando son claves en la catálisis. Otros, sin embargo, presentan mayor divergencia como es el caso de D334 (Ce) y E335 (Hs); o F261 (Ce) y F261 (Hs), por lo que tal vez no tengan un papel tan importante en el proceso.

### 5.3. Comparación de HsPOFUT1 en forma apo y con GDP-fucosa.

Para comparar estas dos conformaciones de la proteína se ha hecho un alineamiento de sus estructuras estabilizadas (FIGURA 16A) para tratar de encontrar diferencias.



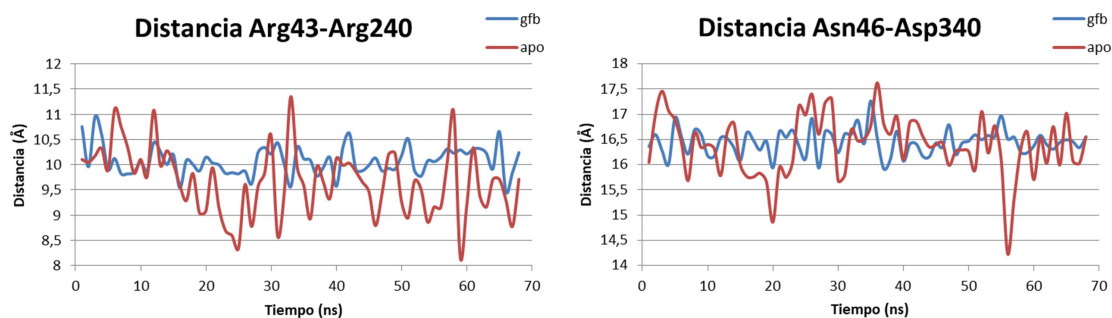


**FIGURA 16.** A: Alineamiento en PyMol de HsPOFUT1 apo (azul) y HsPOFUT1 con ligando (rosa). B: Superposición del centro activo de la HsPOFUT1 apo (azul) y con ligando (rosa). Se muestran las estructuras de los aminoácidos que lo conforman y GDP-fucosa en blanco.

La parte más relevante de esta superposición es la disposición de los aminoácidos que constituyen el centro activo, que se ve ampliada en la FIGURA 16B. Se aprecia una forma más ordenada de los aminoácidos en presencia de GDP-fucosa, con sus cadenas laterales estableciendo las interacciones indicadas con el ligando, mientras que en la forma apo estos aminoácidos se distribuyen de acuerdo al espacio vacío disponible.

Una herramienta que resulta también útil para este objetivo es el estudio de cómo varía la distancia entre dos residuos presentes en el sitio activo de la enzima, en el caso sin el ligando y con el ligando. Estas distancias se pueden calcular para todos los frames de igual forma que el RMSD, y representar gráficamente para ver su evolución a lo largo del proceso.

He escogido como ejemplo el cálculo de las distancias entre los aminoácidos Arg43-Arg240 y Asn46-Asp340, ya que son residuos que se encuentran en lados opuestos del centro activo y pueden dar una idea de la estabilidad de su estructura (FIGURA 17).



**FIGURA 17.** Distancia entre los aminoácidos Arg43-Arg240 (izda.) y Asn46-Asp340 (dcha.) en la HsPOFUT1 en forma apo (rojo) y en HsPOFUT1 con el ligando GDP-fucosa (azul).

Se puede determinar que, aunque en las dos conformaciones de la proteína estas distancias calculadas oscilen alrededor del mismo valor aproximadamente, esta oscilación tiene más amplitud en la forma apo, porque la ausencia de ligando permite mayor libertad de movimiento a estos residuos; mientras que en presencia de GDP-fucosa las distancias tienen menos varianza por interacciones de estabilización con el ligando.

#### 5.4. Estudio de interacciones con el ligando.

En este caso el objetivo es hacer un estudio comparativo entre las interacciones con el ligando predichas mediante modelos computacionales, y las interacciones que se han encontrado empíricamente con las correspondencias halladas en la proteína humana.

Mediante el software por el software de modelado molecular Schrödinger/Maestro<sup>23</sup> se puede confeccionar un mapa 2D de interacciones proteína-ligando (FIGURA 18) que resulta apropiado para este fin:

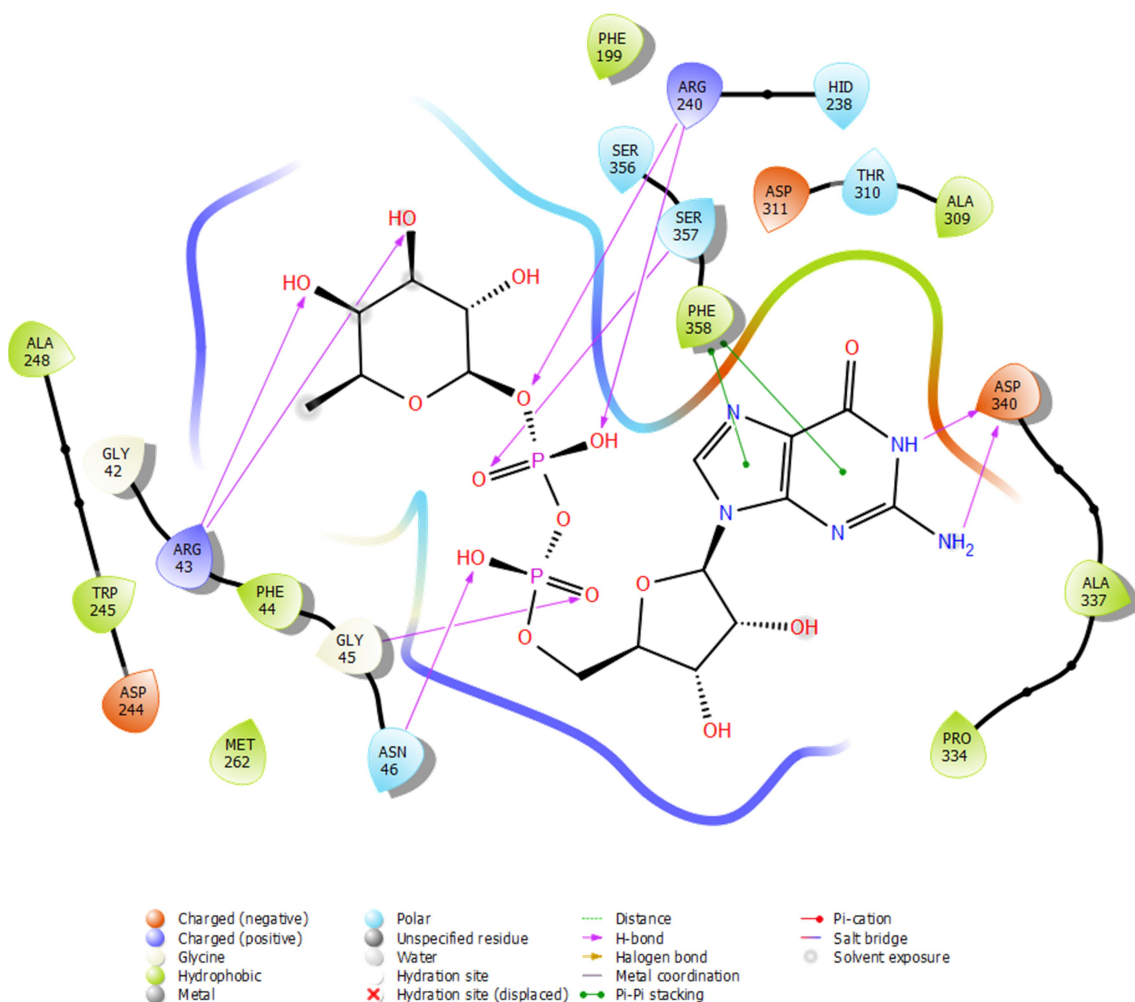


FIGURA 18. Diagrama de interacciones de la proteína HsPOFUT1 con el ligando, con un cutoff de 4 Å. Las conexiones rosas indican puentes de hidrógeno y las verdes interacciones de stacking.

A continuación se resumen las interacciones empíricas con la POFUT1 humana, a partir de las equivalencias halladas con la de *C. elegans* según el trabajo de Lira-Navarrete et al.<sup>13</sup>

- **Con la guanina:**
  - **Phe358:** interacción hidrofóbica de stacking.
  - **Asp311:** interacción hidrofóbica de stacking.
  - **Thr310:** puente de hidrógeno con backbone.
  - **Glu335:** puente de hidrógeno con backbone y cadena lateral.
  - **His238:** puente de hidrógeno con cadena lateral.
- **Con la ribosa:**
  - **Arg43:** puente de hidrógeno con cadena lateral.
  - **Phe44:** puente de hidrógeno con backbone.
- **Con el fosfato- $\alpha$ :**
  - **Gly45:** puente de hidrógeno con backbone.
  - **Asn46:** puente de hidrógeno con backbone.
- **Con el fosfato- $\beta$ :**
  - **Ser356:** puente de hidrógeno con cadena lateral.
  - **Ser357:** puente de hidrógeno con cadena lateral.
  - **Arg240:** puente de hidrógeno y puente salino con cadena lateral.
- **Con la fucosa:**
  - **Phe261:** interacción hidrofóbica de stacking.
  - **Arg240:** puente de hidrógeno con cadena lateral.
  - **Asn46:** puente de hidrógeno con cadena lateral.

En comparación, el software ha sido capaz de predecir varias de estas interacciones de una forma que se aproxima a los resultados experimentales: el stacking de Phe358 y unos cuantos puentes de hidrógeno (Arg43, Gly45, Asn46, Arg240, Asp340 y Ser357).

Esto demuestra la eficacia de los métodos computacionales para obtener información de interacciones de la proteína con el ligando, y trabajar en el diseño de nuevos experimentos utilizando estas predicciones.

## 6. Conclusiones.

En este trabajo se ha procedido al desarrollo de simulaciones de dinámica molecular de tres proteínas POFUT1: la forma apo del organismo *C. elegans*, la apoproteína de *H. sapiens*, y la forma en complejo con su ligando GDP-fucosa también en *H. sapiens*.

La visualización de estas simulaciones, junto con el posterior cálculo de parámetros, como distancias o RMSD, constituye una herramienta fundamental para comprender el proceso de estabilización de estas proteínas.

Se ha hecho un análisis comparativo de las apoproteínas de *C. elegans* y *H. sapiens*, comprobando la conservación de la naturaleza química de los residuos de aminoácidos del centro activo en su interacción con el ligando.

Se han comparado las estructuras de la POFUT1 humana con y sin el ligando, evidenciando el papel estabilizante en los aminoácidos del centro activo por la presencia de la GDP-fucosa.

Se ha utilizado software de modelado molecular para hacer una predicción de las interacciones de la proteína de *H. sapiens* con el ligando, que se demuestra que concuerda con la información obtenida experimentalmente, demostrando la importancia de las herramientas bioinformáticas como base para el estudio de estructuras de proteínas.

## Referencias.

- <sup>1</sup> Rini, J. "Chapter 5: Glycosyltransferases and Glycan-processing enzymes. Essentials of Glycobiology. Varki A, Cummings RD, Esko JD, et al., editors." (2009).
- <sup>2</sup> Ünligil, Uluğ M., and James M. Rini. "Glycosyltransferase structure and mechanism." *Current opinion in structural biology* 10, no. 5 (2000): 510-517.
- <sup>3</sup> McMillan, Brian J., Brandon Zimmerman, Emily D. Egan, Michael Lofgren, Xiang Xu, Anthony Hesser, and Stephen C. Blacklow. "Structure of human POFUT1, its requirement in ligand-independent oncogenic Notch signaling, and functional effects of Dowling-Degos mutations." *Glycobiology* 27, no. 8 (2017): 777-786.
- <sup>4</sup> Wang, Michael M. "Notch signaling and Notch signaling modifiers." *The international journal of biochemistry & cell biology* 43, no. 11 (2011): 1550-1562.
- <sup>5</sup> Haltiwanger, Robert S., and John B. Lowe. "Role of glycosylation in development." *Annual review of biochemistry* 73, no. 1 (2004): 491-537.
- <sup>6</sup> Haltom, Amanda R., and Hamed Jafar-Nejad. "The multiple roles of epidermal growth factor repeat O-glycans in animal development." *Glycobiology* 25, no. 10 (2015): 1027-1042.
- <sup>7</sup> Yin, Ling, Omaid C. Velazquez, and Zhao-Jun Liu. "Notch signaling: emerging molecular targets for cancer therapy." *Biochemical pharmacology* 80, no. 5 (2010): 690-701.
- <sup>8</sup> Steinbuck, Martin Peter, and Susan Winandy. "A review of notch processing with new insights into ligand-independent notch signaling in T-cells." *Frontiers in immunology* 9 (2018): 1230.
- <sup>9</sup> Chabanaïs, Julien, François Labrousse, Alain Chaunavel, Agnès Germot, and Abderrahman Maftah. "POFUT1 as a promising novel biomarker of colorectal cancer." *Cancers* 10, no. 11 (2018): 411.
- <sup>10</sup> Wikipedia, the free encyclopedia. "Notch Signaling Pathway," April 6, 2021. [https://en.wikipedia.org/wiki/Notch\\_signaling\\_pathway](https://en.wikipedia.org/wiki/Notch_signaling_pathway).
- <sup>11</sup> Merino, Pedro, Tomás Tejero, Ignacio Delso, Ramon Hurtado-Guerrero, Asier Gómez-SanJuan, and David Sádaba. "Recent progress on fucosyltransferase inhibitors." *Mini reviews in medicinal chemistry* 12, no. 14 (2012): 1455-1464.
- <sup>12</sup> Bioorganic Chemistry Research Group. "Research." Accessed May 13, 2021. <http://bioorganica.unizar.es/research.html>.
- <sup>13</sup> Lira-Navarrete, Erandi, Jessika Valero-González, Raquel Villanueva, Marta Martínez-Júlvez, Tomás Tejero, Pedro Merino, Santosh Panjekar, and Ramon Hurtado-Guerrero. "Structural insights into the mechanism of protein O-fucosylation." *PLoS one* 6, no. 9 (2011): e25365.
- <sup>14</sup> RCSB Protein Data Bank. "Crystal structure of POFUT1 apo-form (crystal-form-I)." Accessed May 14, 2021. <https://www.rcsb.org/structure/3ZY4>.
- <sup>15</sup> RCSB Protein Data Bank. "Crystal structure of POFUT1 in complex with GDP-fucose (crystal-form-I)." Accessed May 14, 2021. <https://www.rcsb.org/structure/3ZY5>.
- <sup>16</sup> RCSB Protein Data Bank. "Structure of Human POFUT1 in its apo form." Accessed May 21, 2021. <https://www.rcsb.org/structure/5UX6>.
- <sup>17</sup> RCSB Protein Data Bank. "Structure of Human POFUT1 in complex with GDP-fucose." Accessed May 21, 2021. <https://www.rcsb.org/structure/5UXH>.
- <sup>18</sup> D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.

---

<sup>19</sup> UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.

<sup>20</sup> The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

<sup>21</sup> *The Amber residue naming convention for the various histidine connectivities*. BioChemCoRe 2018. "Preparing Your System for Molecular Dynamics (MD)." Accessed May 27, 2021. <https://ctlee.github.io/BioChemCoRe-2018/system-prep/>.

<sup>22</sup> Expasy. "SIM Alignment Tool - Protein Sequences." Accessed June 21, 2021. <https://web.expasy.org/sim/>.

<sup>23</sup> Schrödinger Release 2021-2: Maestro, Schrödinger, LLC, New York, NY, 2021.